Zhaoyang Chu

Tel: +86 15907163884 | chuzhaoyang@hust.edu.cn | zhaoyang-chu.github.io

EDUCATION

Huazhong University of Science and Technology (Advised by Prof. Yao Wan) GPA: 3.53/4.0
M.E. in Computer Science and Technology *Sep. 2022 – Jun. 2025*

Huazhong Agricultural University GPA: **3.93**/4.0
*B.E. in Data Science and Big Data Technology (**Graduated with Honors**)* *Sep. 2018 – Jun. 2022*

RESEARCH INTERESTS

Code Intelligence, e.g., code generation, code search, and vulnerability detection.

Trustworthy Artificial Intelligence, e.g., explainability, privacy, and robustness.

Deep Learning on Graphs, e.g., graph neural networks (GNNs).

PUBLICATIONS

[1] Graph Neural Networks for Vulnerability Detection: A Counterfactual Explanation,
Zhaoyang Chu, Yao Wan*, Qian Li, Yang Wu, Hongyu Zhang, Yulei Sui, Guandong Xu, Hai Jin,
The ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA 2024**).

[2] Hierarchical Graph Representation Learning for the Prediction of Drug-Target Binding Affinity,
Zhaoyang Chu, Feng Huang, Haitao Fu, Yuan Quan, Xionghui Zhou, Shichao Liu, Wen Zhang*,
Information Sciences (**Impact Factor 8.1**), 2022.

RESEARCH PROJECTS

Machine Unlearning for Code Language Models (CLMs) Sep. 2023 – Present
- Propose to erase specific memorization of CLMs to comply with the *Right To Be Forgotten* (RTBF) laws.
- Develop a gradient-based *machine unlearning* approach to make the CLMs forget target information quickly without requiring retraining them from scratch.
- Adopt a divergence-based constraint strategy to minimize the influence of unlearning on the CLM's utility.
- **Submitted to ICSE 2025 (First Author)**, co-advised by Prof. Zhikun Zhang at Stanford.

Learning-based Pre-trained Code Model Selection for Reuse Sep. 2023 – Present
- Investigate the problem of efficiently selecting and reusing pre-trained code models for target tasks.
- Find intuitive strategies (e.g., selecting by model size, by pre-training data, or by brute-force fine-tuning) are not suitable for model selection due to inaccuracy and high computational cost.
- Propose learning-based approaches for accurate model selection within a limited computational budget.
- **Submitted to TSE (Co-Author)**, collaborated with Prof. Hai Jin at Huazhong University of Science and Technology.

Exploring LLMs as Evaluator for Code Summarization Sep. 2023 – Present
- Explore LLMs for assessing code summarization without references.
- Prompt LLMs to adopt diverse roles in evaluating generated summaries across multiple dimensions.
- Find LLM-based evaluator significantly improves the correlation with human perception.
- **Submitted to ICSE 2025 (Co-Author)**, collaborated with Prof. Yulei Sui at the University of New South Wales.

Counterfactual Reasoning for GNN-based Vulnerability Detection Apr. 2023 – Dec. 2023
- Reformulate the problem of explainability in vulnerability detection from a *what-if* analysis perspective.
- Generate counterfactual explanations by identifying minimal perturbations to the input code graph that would alter the decision of the GNN-based detector, thus discovering the root causes of the vulnerability.

- Employ an *edge mask learning* technique for efficient and precise determination of minimal counterfactual perturbations for code graphs.
- **Published at ISSTA 2024 (First Author)**, co-advised by Prof. Qian Li at Curtin University and Prof. Hongyu Zhang at Chongqing University.

GNNs for Drug-Target Binding Affinity (DTA) Prediction   Sep. 2020 – Sep. 2022
- Build a *hierarchical GNN model* to integrate the coarse- and fine-level information from an affinity graph and drug/target molecule graphs, respectively, in a well-designed coarse-to-fine manner.
- Design a similarity-based representation inference method to complement the generalization of the model on predicting affinity values for the new drugs or targets.
- **Published at Information Sciences (First Author)**, advised by Prof. Wen Zhang at Huazhong Agricultural University.

Open-Source Toolkit

NaturalCC: An Open-Source Toolkit for Code Intelligence   Sep. 2022 – Present
- **Main Contributor:** Responsible for compatibility with Transformers and supporting popular large code models like Code Llama, CodeT5, CodeGen, and StarCoder from Hugging Face.

Honors & Awards

Merit Postgraduate, First-class Scholarship for Postgraduates	2022 – 2023
Undergraduate Thesis Innovation Award, Outstanding Undergraduate	2021 – 2022
National Scholarship for Undergraduates, Merit Student	2019 – 2020

Teaching

Teaching Assistant at Huazhong University of Science and Technology
Graph Neural Networks	Sep. 2022 – Dec. 2023
Compiler Principles and Techniques	Mar. 2023 – Jun. 2023

Technical Skills

Programming Languages: Python (Pytorch and Tensorflow), C/C++, Java, SQL, and Scala
Tools: Git, Linux, LaTeX, Tree-sitter, and Joern